Module 9 Lab Activity: Essential Concepts in Multiple Regression
PSY 652 Research Methods
Oct 28, 2020

Description of the dataset:

A research team sought to examine factors associated with 21st birthday drinking among female students at a large University. Female students who were nearing age 21 and self-classified as regular drinkers were eligible for the study. In total, 200 students were recruited and agreed to take part in the study. Students were instructed to report to the lab two weeks prior to their 21st birthday. During this lab session, students completed a brief survey that measured alcohol use during the past month (using the Timeline Follow Back Method) and their weight was recorded. One week prior to their 21st birthday, participants were sent a link for an online survey to measure positive alcohol expectancies for drinking on their 21st birthday. Within three days prior to their 21st birthday, students reported to the lab and were given a diary-based data collection form to record several items on their 21st birthday. Students were instructed to record the food that they consumed during the day, the degree to which they were in a partying mood just prior to the celebration, and the quantity and type of drinks that they consumed during the first two hours of the celebration. The students were also given a small breathalyzer machine to measure BAC 2 hours after consumption of their first drink.
The dataset called bac_obs.csv contains the data:

- *weight: weight in kilograms*
- *alcexp: positive alcohol expectancy for drinking on the impending 21st birthday, a multi-item scale that ranges from 1-7, where a higher score indicates more positive expectations about the role alcohol will play*
- *typ_drks: the number of standard alcohol drinks consumed in the past 30 days*
- *pmood: a rating on a scale from 1-9 on the respondent's mood to party on the 21st birthday, where 1 means never been less in the mood to party, and 9 means never been more in the mood to party*
- *absorb: a score calculated from the food diaries to determine how full the participant was when they began drinking, the score rang- es from 1 to 8, where 1 means a completely full stomach, and 8 means a completely empty stomach*
- *alc_gm: a score calculated from the drinking diary to estimate the grams of alcohol consumed on the 21st birthday*
- *bac: the participant's blood alcohol content, measured as grams of alcohol per deciliter of blood on the 21st birthday*

1. Download the "bac_obs.csv" dataset from the Module 9 Lab and save it into a project folder.
2. Create a new R notebook from your project file and name it "bac_notebook"
3. Create a new R chunk with a first level header: "Load Libraries"
   a. load the psych and tidyverse packages
4. Create a new R chunk with a first level header: "Import Data"
   a. Read in the "bac_obs.csv" dataset, assign it to an object named "bac_obs"
5. Create a new R chunk with a first level header: "Mutate BAC variable"
   a. In this chunk create a new version of the bac variable called bac100 (bac100 = bac*100) and save this to the bac dataframe.
      i. We start here because bac is a very small number and multiplying it by a constant won't change the overall model but will make our output easier to read. (You'll get extravagantly small numbers in scientific notation if you don't do this).
6. Create a new chunk with a first level header: "Select variables of interest"
   a. In an R-chunk create an object called "bac_reduced" and select the following variables to be in the new dataset object: **bac100, weight, alc_gm, typ_drks, alcexp** (There should be 5 variables in your bac_reduced dataset)
7. Create a new R chunk with a first level header: "Describe data"
   a. Use any method to obtain descriptives for the bac_reduced dataset
8. Create a new R chunk with a first level header: "correlate data"
   a. Use any method to get a correlation matrix of the bac_reduced dataset
9. Write a first level header: "Fit regression models"
   a. Create a second level header called: "Regress bac100 on weight and typ_drks"
      i. In a new code chunk. Regress bac on weight and typ_drks. Ask for the output (via summary()).
      ii. Write sentences in the white space to interpret the following: The intercept, the estimates, and p values.
   b. Create a second level header called: "Regress bac on alc_gm and typ_drks"
      i. Add a new code chunk. Regress bac100 on alc_gm and typ_drks. Ask for the output (via summary()).
      ii. Write sentences in the white space to interpret the following: the intercept, the estimates, and p values.
   c. Create a second level header called: "Regress bac on weight, alc_gm, typ_drks and alcexp"
      i. Add a new code chunk. Regress bac100 on weight, alc_gm, typ_drks and alcexp. Ask for the output (via summary()).
      ii. Write sentences in the white space to interpret the following: the intercept, the estimates, and p values.
   d. Examine the output for these three models and answer the following questions in the white space:
      i. How do the R^2 values vary across the three models? Why does the MLR with four predictors have the highest R^2?

ii.   Using your results from model 1 (bac100 regressed on weight & typ_drks) provide me the estimated bac100 for an individual that weighs 70kg and has consumed 27 drinks in the past 30 days.

1. *Note: Divide your result by 100 and you will have the estimated bac level!*